Subspace interference in statistical signal detection

F. C. Nicolls and G. de Jager

Abstract— **Signal detection in certain noise environments fits naturally into a statistical hypothesis testing framework. In order to have moderately tractable models, the noise is often assumed to be additive with a multivariate normal distribution. Additionally, computational complexity requirements may demand the assumption of spatial stationarity, particularly in the case when the data is 2-dimensional.**

Naturally the success of such models is dependent on the degree to which the assumptions are valid. For example, the assumption of spatial stationarity is thrown into serious doubt in typical images, where actual structure exists which might not be amenable to a simple statistical characterisation. In an attempt to improve the validity, we make use of the notion of subspace interference. This assumes that there is an additional unknown signal component present in the data, which is required to lie in a low-dimensional subspace of the original observation space. Invariant hypothesis tests can then be formulated for this problem, which are optimal in a fairly powerful sense.

The work to be presented outlines some of the theory involved in specifying the tests and estimating the model parameters from actual data. Since the emphasis is on image rather than signal processing, special care needs to be taken in developing computational methods for the solutions: simple-minded approaches have massive computation and memory requirements.

I. Introduction

A very natural way of dealing with the problem of detection of transient targets in a signal is to use a sliding window approach. This has the advantage of simplicity and computational appeal, although it does require the assumption that detection can be regarded as a localised operation.

The strength of the sliding window method is hugely enhanced if the assumption of spatial stationarity can be made. This allows for a tractable and simple signal model for which a detector can be implemented with minimal difficulty. The advantages are so great that it is questionable whether the sliding window method should be used at all if this assumption is significantly violated. In that case, time-frequency or time-scale representations are probably more useful.

In some situations the assumption of stationarity is compromised by certain interferences in the signal. For example, a signal may be contaminated by a DC offset or by an approximately linear trend which adds to it. Images, too, often contain regions which are very smooth, but which are at an indeterminate intensity and which may vary slowly with changes in spatial position. Such interferences may be modelled as subspace interferences, because they are constrained to lie in some linear subspace of the space of observations. They are sufficient to compromise the stationary assumption.

It would be a shame to have to abandon the sliding window approach because of these relatively simple interferences. Fortunately, the notion of invariance in hypothesis testing can be used to specify detectors for these situations which conform to a fairly natural optimality criterion. The design and implementation of such detectors, particularly for the case of images, was the subject of a previous technical report [5].

In that report, it was assumed that the interference subspace is known, and that the nominal distribution of the data in the window (before interference is added) is multivariate normal (MVN) with known parameters. In real situations, however, the interference subspace might be unknown and the model parameters will probably have to be estimated from a sample of target-free data. The target to be detected will usually also have to play a role in the specification of these parameters, since at best any given detector can only be optimal for a restricted class of signals.

This report assumes that the interference subspace has been specified in advance, and discusses issues related to estimating the covariance matrix model parameters. For simplicity, only the case of 1-D signals will be dealt with, although with some effort the results can be extended to 2-D.

II. Overview of detection in coloured noise

Subspace interference is a particular type of interference where every observation has an unwanted additive contribution which is constrained to lie in some linear subspace of the original observation space. Thus if the underlying random process is distributed according to $\mathbf{x} : N[\mathbf{m}, \mathbf{T}]$ and the independent realisations of this process are labelled $\mathbf{x}_1, \ldots, \mathbf{x}_m$, then the observed data are $\mathbf{y}_1, \ldots, \mathbf{y}_m$ with

$$\mathbf{y}_j = \mathbf{x}_j + \mathbf{U}_I \mathbf{c}_j. \quad (1)$$

Digital Image Processing, Department of Electrical Engineering, University of Cape Town, Rondebosch 7701, South Africa. E-mail: nicolls@dip.ee.uct.ac.za.

In this equation $\mathbf{U}_I$ is a matrix which spans the interference subspace, and for convenience we may assume that $\mathbf{U}_I^T\mathbf{U}_I = \mathbf{I}$. The constant vectors $\mathbf{c}_j$ differ for each observation, and are completely unknown. For definiteness, we assume that the covariance matrix $\mathbf{T}$ has a Toeplitz structure, in accordance with the stationary assumption.

It is possible to formulate a constant false alarm rate hypothesis test which is invariant to such interferences. Some of the theory and computational issues involved will be dealt with here. The two cases of detection with and without an invariance subspace are discussed in this section.

A. Detection without invariance

This is by far the simpler of the two cases: the hypotheses are $\mathbf{x} : N[\mathbf{m}_0, \mathbf{T}]$ under H_0 and $\mathbf{x} : N[\mathbf{m}_1, \mathbf{T}]$ under H_1. The log likelihood ratio for this problem is [7]

$$L(\mathbf{x}) = (\mathbf{m}_1 - \mathbf{m}_0)^T \mathbf{T}^{-1}(\mathbf{x} - \mathbf{x}_0), \tag{2}$$

where $\mathbf{x}_0 = 1/2(\mathbf{m}_1 + \mathbf{m}_0)$. The optimal test is to compare this to a threshold p, and select H_1 when this threshold is exceeded: choose H_1 when

$$(\mathbf{m}_1 - \mathbf{m}_0)^T \mathbf{T}^{-1}(\mathbf{x} - \mathbf{x}_0) > p. \tag{3}$$

Letting $\mathbf{w}^T = (\mathbf{m}_1 - \mathbf{m}_0)^T \mathbf{T}^{-1}$, the decision inequality can be written as

$$\mathbf{w}^T\mathbf{x} > p + \mathbf{w}^T\mathbf{x}_0, \tag{4}$$

which shows that the inner product $\mathbf{w}^T\mathbf{x}$ is sufficient for the decision process. For purposes of implementation in a sliding window framework, this statistic can be calculated (for each signal location) by means of a convolution operation with template $\mathbf{w}$.

It is simple to calculate $\mathbf{w}$ for this problem: since $\mathbf{w}^T = (\mathbf{m}_1 - \mathbf{m}_0)^T \mathbf{T}^{-1}$ it can be seen that

$$\mathbf{T}\mathbf{w} = \mathbf{m}_1 - \mathbf{m}_0, \tag{5}$$

which is just a linear regression with a structured covariance matrix $\mathbf{T}$. Since $\mathbf{T}$ is Toeplitz, $\mathbf{w}$ can be found by standard Levinson-Durbin recursion [2].

B. Detection with invariance

Suppose $\mathbf{U} = (\mathbf{U}_I \;\; \mathbf{U}_H)$ is a unitary matrix, with $\mathbf{U}_I$ spanning the interference subspace to which we require invariance and $\mathbf{U}_H$ spanning the complementary subspace. It was shown in [5] that $\mathbf{y} = \mathbf{U}_H^T\mathbf{x}$ is a maximal invariant statistic for the invariant decision problem.

The distribution of this statistic is $N[\mathbf{U}_H^T\mathbf{m}_0, \mathbf{U}_H^T\mathbf{T}\mathbf{U}_H]$ under H_0 and $N[\mathbf{U}_H^T\mathbf{m}_1, \mathbf{U}_H^T\mathbf{T}\mathbf{U}_H]$ under H_1, so the invariant decision rule is to choose H_1 when

$$(\mathbf{U}_H^T\mathbf{m}_1 - \mathbf{U}_H^T\mathbf{m}_0)^T(\mathbf{U}_H^T\mathbf{T}\mathbf{U}_H)^{-1}(\mathbf{U}_H^T\mathbf{x} - \mathbf{U}_H^T\mathbf{x}_0) > p, \tag{6}$$

or equivalently, when

$$(\mathbf{m}_1 - \mathbf{m}_0)^T\mathbf{U}_H(\mathbf{U}_H^T\mathbf{T}\mathbf{U}_H)^{-1}\mathbf{U}_H^T(\mathbf{x} - \mathbf{x}_0) > p. \tag{7}$$

Taking the same route as for the previous case, we define

$$\mathbf{w}^T = (\mathbf{m}_1 - \mathbf{m}_0)^T\mathbf{U}_H(\mathbf{U}_H^T\mathbf{T}\mathbf{U}_H)^{-1}\mathbf{U}_H^T, \tag{8}$$

and as before the decision becomes choosing H_1 when

$$\mathbf{w}^T\mathbf{x} > p + \mathbf{w}^T\mathbf{x}_0. \tag{9}$$

Using the definition of $\mathbf{w}^T$, it can be seen that

$$\mathbf{w} = \mathbf{U}_H(\mathbf{U}_H^T\mathbf{T}\mathbf{U}_H)^{-1}\mathbf{U}_H^T(\mathbf{m}_1 - \mathbf{m}_0). \tag{10}$$

It is much less obvious how to find the vector $\mathbf{w}$ for this situation. However, some of the methods provided in [5] provide a solution. Partition the matrix $\mathbf{C} = \mathbf{U}^T\mathbf{T}\mathbf{U}$ as

$$\mathbf{C} = \begin{pmatrix} \mathbf{U}_I^T\mathbf{T}\mathbf{U}_I & \mathbf{U}_I^T\mathbf{T}\mathbf{U}_H \\ \mathbf{U}_H^T\mathbf{T}\mathbf{U}_I & \mathbf{U}_H^T\mathbf{T}\mathbf{U}_H \end{pmatrix} = \begin{pmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} \\ \mathbf{C}_{21} & \mathbf{C}_{22} \end{pmatrix} \tag{11}$$

with an inverse (since $\mathbf{U}^{-1} = \mathbf{U}^T$) of

$$\mathbf{A} = \begin{pmatrix} \mathbf{U}_I^T\mathbf{T}^{-1}\mathbf{U}_I & \mathbf{U}_I^T\mathbf{T}^{-1}\mathbf{U}_H \\ \mathbf{U}_H^T\mathbf{T}^{-1}\mathbf{U}_I & \mathbf{U}_H^T\mathbf{T}^{-1}\mathbf{U}_H \end{pmatrix} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix}. \tag{12}$$

Expanding the identity $\mathbf{C}\mathbf{A} = \mathbf{I}$, it can be shown that

$$\mathbf{C}_{22}^{-1} = \mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}, \tag{13}$$

from which the following relation results:

$$(\mathbf{U}_H^T\mathbf{T}\mathbf{U}_H)^{-1} = \mathbf{U}_H^T\mathbf{T}^{-1}\mathbf{U}_H - \mathbf{U}_H^T\mathbf{T}^{-1}\mathbf{U}_I(\mathbf{U}_I^T\mathbf{T}^{-1}\mathbf{U}_I)^{-1}\mathbf{U}_I^T\mathbf{T}^{-1}\mathbf{U}_H. \tag{14}$$

Noting that $\mathbf{U}_H\mathbf{U}_H^T = \mathbf{I} - \mathbf{U}_I\mathbf{U}_I^T$, we can form the product

$$\mathbf{U}_H(\mathbf{U}_H^T\mathbf{T}\mathbf{U}_H)^{-1}\mathbf{U}_H^T = (\mathbf{I} - \mathbf{U}_I\mathbf{U}_I^T)[\mathbf{T}^{-1} - \mathbf{T}^{-1}\mathbf{U}_I(\mathbf{U}_I^T\mathbf{T}^{-1}\mathbf{U}_I)^{-1}\mathbf{U}_I^T\mathbf{T}^{-1}](\mathbf{I} - \mathbf{U}_I\mathbf{U}_I^T), \tag{15}$$

which simplifies to

$$\mathbf{U}_H(\mathbf{U}_H^T\mathbf{T}\mathbf{U}_H)^{-1}\mathbf{U}_H^T = \mathbf{T}^{-1} - \mathbf{T}^{-1}\mathbf{U}_I(\mathbf{U}_I^T\mathbf{T}^{-1}\mathbf{U}_I)^{-1}\mathbf{U}_I^T\mathbf{T}^{-1}. \tag{16}$$

Substituting into equation 10, the vector $\mathbf{w}$ is given by

$$\mathbf{w} = \mathbf{T}^{-1}(\mathbf{m}_1 - \mathbf{m}_0) - \mathbf{T}^{-1}\mathbf{U}_I(\mathbf{U}_I^T\mathbf{T}^{-1}\mathbf{U}_I)^{-1}\mathbf{U}_I^T\mathbf{T}^{-1}(\mathbf{m}_1 - \mathbf{m}_0). \tag{17}$$

Since the Toeplitz system solvers can be used to find both $\mathbf{T}^{-1}(\mathbf{m}_1 - \mathbf{m}_0)$ and the fairly low-dimensional matrix $\mathbf{T}^{-1}\mathbf{U}_I$, this quantity can be calculated without too much difficulty. Note that at no stage is it required to specify $\mathbf{U}_H$ explicitly.

III. Maximum likelihood parameter estimation with subspace interference

In the previous section it was shown that it is possible to design a detector which is invariant to subspace interferences. Once this detector is specified, the actual implementation is no different from that of a simple detector in uncorrelated noise. However, the issue arises of how to find the model parameters to be used in this detector.

It is assumed that the parameters of the random process $\mathbf{X} : N[\mathbf{0}, \mathbf{T}]$ need to be estimated. However, realisations of this process cannot be observed directly; each observation is contaminated by subspace interference. The observed data $\mathbf{y}_1, \ldots, \mathbf{y}_m$ for the independent realisations $\mathbf{x}_1, \ldots, \mathbf{x}_m$ of the process are related to one another by equation 1.

The problem described here is amenable to a missing data interpretation: the component of $\mathbf{X}$ which lies in the interference subspace is effectively destroyed by the unknown contribution $\mathbf{c}_j$, and is therefore useless for inferential purposes. This is easily demonstrated by expanding the observation $\mathbf{y}_j$ onto the orthonormal basis $\mathbf{U} = (\mathbf{U}_I\ \mathbf{U}_H)$, where $\mathbf{U}_H$ spans the subspace complementary to $\mathbf{U}_I$:

$$\begin{aligned}\begin{pmatrix}\mathbf{U}_I & \mathbf{U}_H\end{pmatrix}^T \mathbf{y}_j &= \begin{pmatrix}\mathbf{U}_I^T \\ \mathbf{U}_H^T\end{pmatrix} (\mathbf{x}_j + \mathbf{U}_I \mathbf{c}_j) \\ &= \begin{pmatrix}\mathbf{U}_I^T \mathbf{x}_j \\ \mathbf{U}_H^T \mathbf{x}_j\end{pmatrix} + \begin{pmatrix}\mathbf{c}_j \\ \mathbf{0}\end{pmatrix}. \end{aligned} \qquad (18)$$

Since $\mathbf{c}_j$ is completely unspecified, the first i dimensions of this vector are effectively destroyed by the interference. In the original coordinate system, this means that only the component $\mathbf{y}_j^e = \mathbf{U}_H^T \mathbf{x}_j = \mathbf{U}_H^T \mathbf{y}_j$ of the observation is valid for estimating the parameters in the distribution of $\mathbf{X}$.

It might not be completely evident why it is important to ignore the interference subspace when estimating the structured covariance matrix. The following scenario should make the need evident: suppose the underlying distribution is two-dimensional with an identity covariance matrix. The solid circle in figure 1 shows a contour of constant probability for this case. If there is interference in the direction of $\mathbf{x}_2$, the sample covariance in this direction will be increased. Thus the overall sample covariance might be as represented by the dashed ellipse in that figure. Now since a Toeplitz covariance matrix has the same value along the major diagonal, the variance must be equal along all the axes of the coordinate system. The dotted contour in figure 1 therefore indicates the constrained estimate. Considering only the interference-free subspace $\mathbf{x}_1$, it can be seen that the variance in this dimension has been overestimated due to the interference in $\mathbf{x}_2$.

Returning to the problem, the portion of the data that is uncorrupted by interference is $\mathbf{Y}^e : N[\mathbf{0}, \mathbf{U}_H^T \mathbf{T} \mathbf{U}_H]$. It is

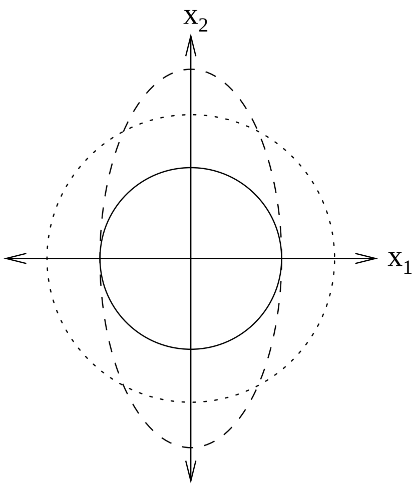

Fig. 1. Contours of constant probability for the actual distribution (solid), the sample covariance (dashed), and the estimated covariance (dotted).

required to estimate $\mathbf{T}$ from m samples of this quantity, under the constraint that $\mathbf{T}$ be Toeplitz. If m is large, maximum likelihood estimation is approximately optimal. This is typical of many signal processing problems in engineering — there is too much data rather than too little. Thus the more subtle estimation procedures are effectively superseded by the simpler likelihood formulations.

The EM algorithm is commonly used for maximum likelihood estimation of Toeplitz covariances. This is an iterative algorithm whereby a difficult parametric optimisation problem is embedded inside a higher-dimensional but computationally more tractable one [1]. It is an ideal formulation for the problem outlined here: the hypothetical u-dimensional complete data observations are $\mathbf{z}_j : N[\mathbf{0}, \mathbf{C}]$, with $\mathbf{C}$ a circulant matrix representing the parameter to be optimised over, and the actual useful observations $\mathbf{y}_j^e$ take the role of the incomplete data. The embedding is such that the interference-free data $\mathbf{x}_j$ (which is never observed) constitutes the first r elements of the u-dimensional complete data samples $\mathbf{z}_j$. Thus

$$\mathbf{x}_j = \mathbf{I}_{r\times u} \mathbf{z}_j, \qquad (19)$$

where $\mathbf{I}_{r\times u}$ is the $r \times u$ identity matrix of zeros with ones along the main diagonal. The useful observations $\mathbf{y}_j^e$ are related to the complete data $\mathbf{z}_j$ by

$$\begin{aligned}\mathbf{y}_j^e &= \mathbf{U}_H^T \mathbf{y}_j = \mathbf{U}_H^T \mathbf{x}_j \\ &= \mathbf{U}_H^T \mathbf{I}_{r\times u} \mathbf{z}_j. \end{aligned} \qquad (20)$$

The reason for the EM algorithm being effective in this problem is because a circulant matrix is very easily diagonalisable. The details of the EM solution to this problem is the topic of the next section.

The method of solution redefines the problem slightly: instead of maximising the likelihood over the set of all Toeplitz matrices, the maximisation is performed over the set of all matrices with positive definite circulant extensions. This is a standard formulation which can be found for example in [4]. The covariance matrix $\mathbf{T}$ can be obtained from the corresponding complete data circulant covariance matrix $\mathbf{C}$ by

$$\mathbf{T} = \mathbf{I}_{r\times u}\mathbf{C}\mathbf{I}_{r\times u}^T. \tag{21}$$

A notable feature of the EM algorithm is its use of a missing data formalism to arrive at the required solution. Since the problem of subspace interference is also conducive to a missing data interpretation, the method is further justified for use in this situation.

IV. EM FORMULATION OF SOLUTION

The quantity $\mathbf{Y}^e = \mathbf{U}_H^T\mathbf{X}$ is all that is observed of the hypothetical u-dimensional complete data $\mathbf{Z} : N[\mathbf{0},\mathbf{C}]$, where $\mathbf{C}$ is circulant. For m independent observations, the complete data probability density function is

$$p_{\mathbf{C}}(\mathbf{z}_1,\dots,\mathbf{z}_m) = (2\pi)^{-mu/2}\log|\mathbf{C}|^{-m/2}e^{-\frac{1}{2}\sum_{j=1}^m \mathbf{z}_j^T\mathbf{C}^{-1}\mathbf{z}_j}. \tag{22}$$

It is simpler to consider the problem in a rotated coordinate system where the covariance matrix is diagonalised.

Let $\mathbf{W}_u$ be the u-dimensional unitary DFT matrix. This matrix diagonalises the class of all $u\times u$ circulant matrices [3, p. 150]. The transformed complete data $\mathbf{D} = \mathbf{W}_u\mathbf{Z}$ is therefore distributed as $\mathbf{D} : N[\mathbf{0},\Sigma]$, with $\Sigma = \mathbf{W}_u\mathbf{C}\mathbf{W}_u^\dagger = \mathrm{diag}[\sigma_1,\dots,\sigma_u]$ a diagonal matrix comprised of the eigenvalues of $\mathbf{C}$. The log-likelihood in this rotated coordinate system is

$$\begin{aligned} L(\Sigma,\mathbf{d}_1,\dots,\mathbf{d}_m) &= K - \frac{m}{2}\log|\Sigma| - \frac{1}{2}\sum_{j=1}^m \mathbf{d}_j^\dagger\Sigma^{-1}\mathbf{d}_j \\ &= K - \frac{m}{2}\sum_{k=1}^u \log\sigma_k - \frac{1}{2}\sum_{k=1}^u\sum_{j=1}^m \frac{|\mathbf{d}_j(k)|^2}{\sigma_k}, \end{aligned} \tag{23}$$

where $\mathbf{d}_j = [\mathbf{d}_j(1),\cdots,\mathbf{d}_j(u)]^T$.

Consider the parameter to be estimated to be the diagonalised covariance matrix Σ, which uniquely specifies the complete data circulant covariance. The EM algorithm proceeds as follows: for the E (expectation) step, the current best estimate $\Sigma^{(p)}$ of the parameter is used to find the expected log-likelihood function $L(\Sigma,\mathbf{d}_1,\dots,\mathbf{d}_m)$, conditioned on the observations $\mathbf{y}_1^e,\dots,\mathbf{y}_m^e$. In the M (maximisation) step, this conditional expectation is maximised with respect to the parameters to yield the next iterate $\Sigma^{(p+1)}$. For the problem addressed in this paper, these steps will now be formalised.

A. Expectation step

Given the previous best estimate $\Sigma^{(p)}$ of the parameter as well as the incomplete data $\mathbf{y}_1^e,\dots,\mathbf{y}_m^e$, the expected value of the complete data log likelihood is

$$\begin{aligned} E\{L|\Sigma^{(p)},\mathbf{y}_1^e,\dots,\mathbf{y}_m^e\} &= K - \frac{m}{2}\sum_{k=1}^u \log\sigma_k^{(p)} - \\ &\frac{1}{2}\sum_{k=1}^u\sum_{j=1}^m \frac{E\{|d_j(k)|^2|\Sigma^{(p)},\mathbf{y}_1^e,\dots,\mathbf{y}_m^e\}}{\sigma_k}. \end{aligned} \tag{24}$$

B. Maximisation step

This involves finding the new parameter $\Sigma^{(p+1)}$ which maximise the conditional expected log-likelihood in equation 24. Taking the derivative with respect to σ_l and setting to zero yields a necessary condition for a maximum ($\forall l$):

$$\begin{aligned} \frac{\partial E\{L|\Sigma^{(p)},\mathbf{y}_1^e,\dots,\mathbf{y}_m^e\}}{\partial\sigma_l} &= -\frac{m}{2}\frac{1}{\sigma_l} - \\ &\sum_{j=1}^m -\frac{E\{|d_j(l)|^2|\Sigma^{(p)},\mathbf{y}_1^e,\dots,\mathbf{y}_m^e\}}{\sigma_l^2} = 0 \\ \Longrightarrow \sigma_l^{(p+1)} &= \frac{1}{m}\sum_{j=1}^m E\{|d_j(l)|^2|\Sigma^{(p)},\mathbf{y}_1^e,\dots,\mathbf{y}_m^e\}. \end{aligned} \tag{25}$$

Given the values of each of the $\sigma_l^{(p+1)}$, the new estimate of the parameter is

$$\Sigma^{(p+1)} = \begin{pmatrix} \sigma_1^{(p+1)} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \sigma_u^{(p+1)} \end{pmatrix}. \tag{26}$$

Since $\mathbf{C}^{(p+1)} = \mathbf{W}_u^\dagger\Sigma^{(p+1)}\mathbf{W}_u$, the improved covariance matrix estimate $\mathbf{T}^{(p+1)}$ can be obtained using equation 21.

V. CALCULATING THE ITERATION

The parameter estimate $\sigma_l^{(p+1)}$ in equation 25 is expressed in terms of the expectations

$$E\{|d_j(l)|^2|\Sigma^{(p)},\mathbf{y}_1^e,\dots,\mathbf{y}_m^e\}, \tag{27}$$

which have yet to be calculated. Taking the same approach as Miller et al. [4], we note that $\sigma_l^{(p+1)}$ in that equation is identical to the lth diagonal element of the matrix

$$\Sigma_{dd}^{(p+1)} = \frac{1}{m}\sum_{j=1}^m E\{\mathbf{d}_j\mathbf{d}_j^\dagger|\Sigma^{(p)},\mathbf{y}_1^e,\dots,\mathbf{y}_m^e\}. \tag{28}$$

Since the observations are independent,

$$E\{\mathbf{d}_j\mathbf{d}_j^\dagger|\Sigma^{(p)},\mathbf{y}_1^e,\ldots,\mathbf{y}_m^e\} = E\{\mathbf{d}_j\mathbf{d}_j^\dagger|\Sigma^{(p)},\mathbf{y}_j^e\}. \qquad (29)$$

To calculate this expectation, the joint distribution of $\mathbf{d}_j$ and $\mathbf{y}_j^e$ is required: with $\mathbf{K}_{yy} = \mathbf{U}_H^T\mathbf{T}^{(p)}\mathbf{U}_H$ and $\mathbf{K}_{dd} = \Sigma^{(p)}$ we have

$$\begin{pmatrix}\mathbf{y}_j^e \\ \mathbf{d}_j\end{pmatrix} \Big| \Sigma^{(p)} : N\left[\begin{pmatrix}\mathbf{0} \\ \mathbf{0}\end{pmatrix}, \begin{pmatrix}\mathbf{K}_{yy} & \mathbf{K}_{yd} \\ \mathbf{K}_{dy} & \mathbf{K}_{dd}\end{pmatrix}\right]. \qquad (30)$$

The expectation $\mathbf{K}_{dy} = \mathbf{K}_{yd}^\dagger$ can be calculated as follows:

$$\begin{aligned}\mathbf{K}_{dy} &= E\{\mathbf{d}_j\mathbf{y}_j^{e\dagger}\} \\ &= E\{\mathbf{d}_j\mathbf{z}_j^\dagger\}\mathbf{I}_{r\times u}^T\mathbf{U}_H \\ &= E\{\mathbf{d}_j\mathbf{d}_j^\dagger\}\mathbf{W}_u\mathbf{I}_{u\times r}\mathbf{U}_H \\ &= \Sigma^{(p)}\mathbf{W}_u^{(r)}\mathbf{U}_H, \end{aligned} \qquad (31)$$

where $\mathbf{W}_u^{(r)}$ contains the first r columns of $\mathbf{W}_u$. The conditional distribution of $\mathbf{d}_j$ given $\mathbf{y}_j^e$ and $\Sigma^{(p)}$ is [6]

$$\mathbf{d}_j|\Sigma^{(p)},\mathbf{y}_j^e : N[\mathbf{K}_{dy}\mathbf{K}_{yy}^{-1}\mathbf{y}_j^e, \Sigma^{(p)} - \mathbf{K}_{dy}\mathbf{K}_{yy}^{-1}\mathbf{K}_{yd}], \qquad (32)$$

from which it can be shown that

$$E\{\mathbf{d}_j\mathbf{d}_j^\dagger|\Sigma^{(p)},\mathbf{y}_j^e\} = \mathbf{K}_{dy}\mathbf{K}_{yy}^{-1}\mathbf{y}_j^e\mathbf{y}_j^{e\dagger}\mathbf{K}_{yy}^{-1}\mathbf{K}_{yd} + \Sigma^{(p)} - \mathbf{K}_{dy}\mathbf{K}_{yy}^{-1}\mathbf{K}_{yd}. \qquad (33)$$

Using this result along with $\mathbf{y}_j^e = \mathbf{U}_H^T\mathbf{y}_j$ in equation 28 yields

$$\Sigma_{dd}^{(p+1)} = \mathbf{K}_{dy}\mathbf{K}_{yy}^{-1}\mathbf{U}_H^T\mathbf{S}_{yy}\mathbf{U}_H\mathbf{K}_{yy}^{-1}\mathbf{K}_{yd} + \Sigma^{(p)} - \mathbf{K}_{dy}\mathbf{K}_{yy}^{-1}\mathbf{K}_{yd}, \qquad (34)$$

where $\mathbf{S}_{yy} = \frac{1}{m}\sum_{j=1}^m \mathbf{y}_j\mathbf{y}_j^\dagger$ is the sample covariance of the actual observation. Writing explicitly, the parameters $\sigma_l^{(p+1)}$ are the diagonal elements of

$$\begin{aligned}\Sigma_{dd}^{(p+1)} = {} & \Sigma^{(p)}\mathbf{W}_u^{(r)}\mathbf{U}_H(\mathbf{U}_H^T\mathbf{T}^{(p)}\mathbf{U}_H)^{-1}\mathbf{U}_H^T\mathbf{S}_{yy}\mathbf{U}_H \\ & (\mathbf{U}_H^T\mathbf{T}^{(p)}\mathbf{U}_H)^{-1}\mathbf{U}_H^T\mathbf{W}_u^{(r)\dagger}\Sigma^{(p)} + \Sigma^{(p)} - \\ & \Sigma^{(p)}\mathbf{W}_U^{(r)}\mathbf{U}_H(\mathbf{U}_H^T\mathbf{T}^{(p)}\mathbf{U}_H)^{-1}\mathbf{U}_H^T\mathbf{W}_u^{(r)^\dagger}\Sigma^{(p)}. \end{aligned} \qquad (35)$$

With inventive use of discrete Fourier transforms and Toeplitz system solvers [4], it is possible to calculate the required elements even for very large matrices $\mathbf{T}$.

VI. Conclusions

Optimal detectors have been derived for the problem of detecting a signal in environments of just noise as well as noise with subspace interference. Efficient methods of calculating the parameters for these detectors have been developed for the case where the covariance matrix is Toeplitz, corresponding to the assumption of spatial stationarity.

The question of how to estimate the covariance matrix parameters when the observations are subject to subspace interference has also been discussed. A modified EM iteration is derived which ignores data contributions which lie within a known linear subspace.

References

[1] A. P. Dempster, N. M. Laird, and D. B. Rubin, “Maximum likelihood from incomplete data via the EM algorithm,” *Journal of the Royal Statistical Society: B*, vol. 39, pp. 1–38, 1977.

[2] G. H. Golub and C. F. Van Loan, *Matrix Computations*. Johns Hopkins University Press, 1983.

[3] A. K. Jain, *Fundamentals of Digital Image Processing*. Prentice-Hall, 1989.

[4] M. I. Miller, D. R. Fuhrmann, J. A. O'Sullivan, and D. L. Snyder, “Maximum-likelihood methods for Toeplitz covariance estimation and radar imaging,” in *Advances in Spectrum Analysis and Array Processing* (S. Haykin, ed.), vol. II, pp. 145–172, Prentice-Hall, 1991.

[5] F. C. Nicolls, “Statistical methods for 2-D signals,” tech. rep., Unviersity of Cape Town, Department of Electrical Engineering, Rondebosch, South Africa, 7701, June 1997.

[6] I. B. Rhodes, “A tutorial introduction to estimation and filtering,” *IEEE Transactions on Automatic Control*, vol. AC-16, pp. 688–706, Dec. 1971.

[7] L. L. Scharf, *Statistical Signal Processing: Detection, Estimation, and Time Series Analysis*. Addison-Weseley Series in Electrical and Computer Engineering: Digital Signal Processing, Addison-Wesley, 1991.